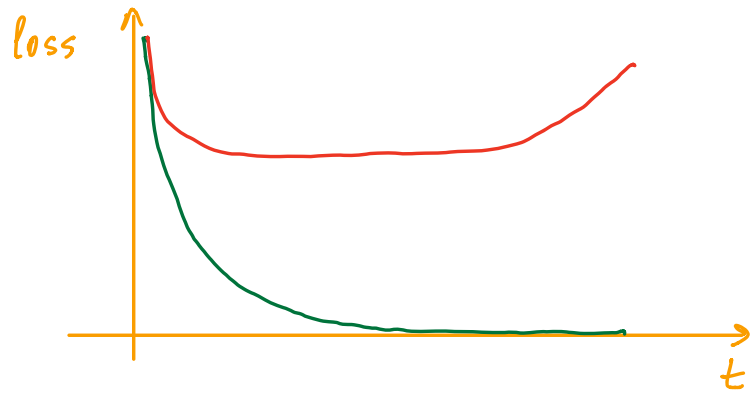
$x \sim X$ | SGD |

1). $b \sim 2^X$; $d \sim B(p)$.

2) $W = W - \lambda \cdot \nabla_W L(W, b)$.

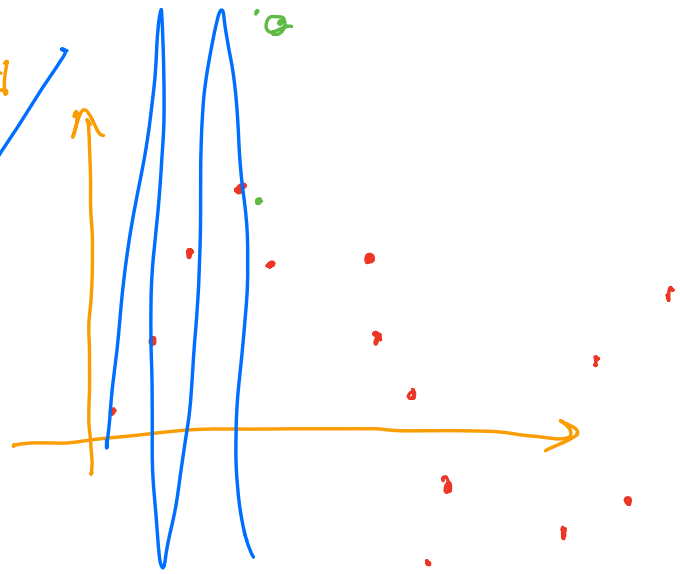


val.

train.

$\tilde{L} = L + P \cdot R$

$R = \|W\|$



$h_1, 512.$

$h_1 = \psi(W_1^T x) \cdot d_1 \cdot \frac{1}{P}$ $\quad h_i \in R^{512}$

512×512

$h_2 = \psi(W_2^T h_1) \cdot d_2 \cdot \frac{1}{P}$ $\quad \hat{h}_i = h_i \cdot d \cdot \frac{1}{P}$

$\vdots$

$d \sim B(P)^{512}$

$h_n = \psi(W_n^T h_n) \cdot d_n \cdot \frac{1}{P}$ $\quad d \in R^{512}$

$\begin{cases} 1, & P \\ 0, & 1-P \end{cases}$